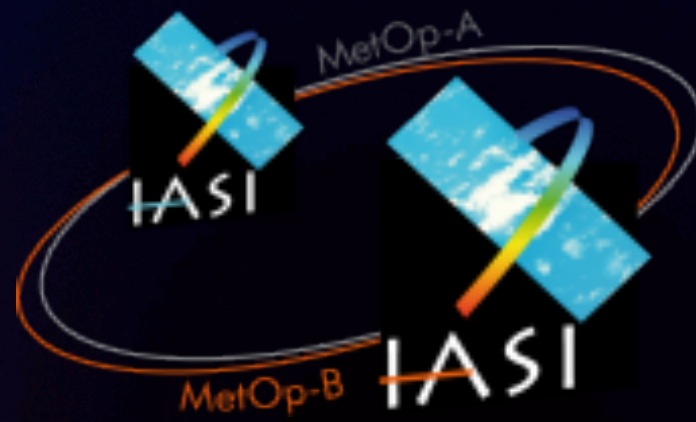
Piecewise linear regression (PWLR) and optimal estimation (OE) retrieval of temperature and humidity in EUMETSAT's IASI Level 2 Processor.

EUMETSAT

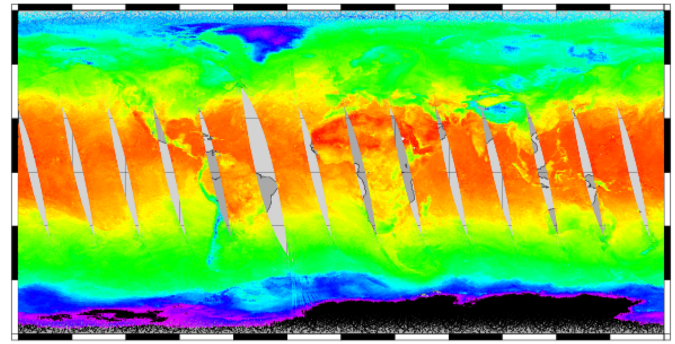
TIM HULTBERG, THOMAS AUGUST



Outline (EUMESAT IASI L2 Version 6 T, W and O profiles)

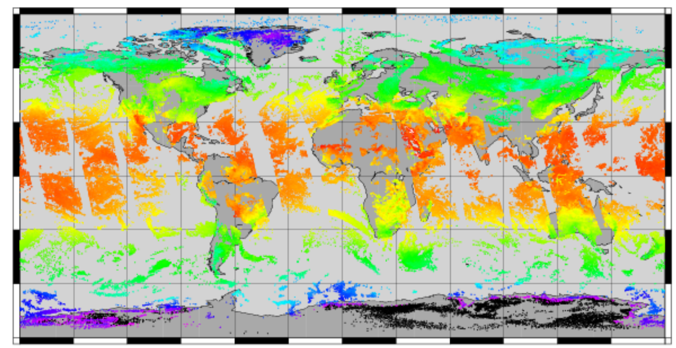
A. First guess / a-priori all sky

- Piecewise linear regression (PWLR) from IASI, AMSU and MHS
- Quality indicators (QI)



B. Optimal estimation in clear sky

- Reconstructed radiances and channel selection
- Signal and forward model subspaces (removal of instrument artefacts)

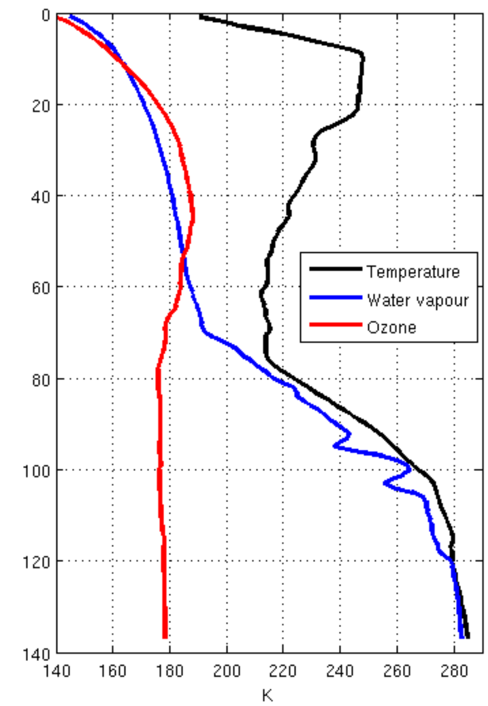


MW + IR Piecewise linear regression

All sky, piece wise linear regression retrievals of temperature, humidity and ozone profiles from co-located IASI, AMSU and MHS measurements.

111 predictors:

- Surface height (km)
- Secant of satellite zenith angle
- Radiance in 14 AMSU channels (channel 7 excluded)
- Radiance in 5 MHS channels
- 30 leading IASI band 1 PC scores
- 30 leading IASI band 2 PC scores
- 30 leading IASI band 3 PC scores



415 dependent variables:

- Ta (K)
- Wa (K)
- Ts (K)
- Sp (hPa)
- T profile (K) at 137 model levels
- W profile (K) dew point temperature at 137 model levels
- O profile (K) "dew point temperature" (W formula) at 137 model levels

Training with real measurements and ECMWF

Training data based on co-located ECMWF analysis from 23 days (1st and 17th of each month from July 2013 to June 2014)

64 regression classes for which individual regression coefficients are retrieved. The class is determined by the land fraction, the surface height as well as the AMSU and MHS radiances.

Class A: LF==0 and AMSU_4 > 370 + 1.5*AMSU_2 (open sea)
Class B: LF==0 and AMSU_4 <= 370 + 1.5*AMSU_2 (sea ice, some clouds over sea)
Class C: LF>0 and z < 825.5 (low elevation land)
Class C: LF>0 and z > 825.5 (high elevation land)

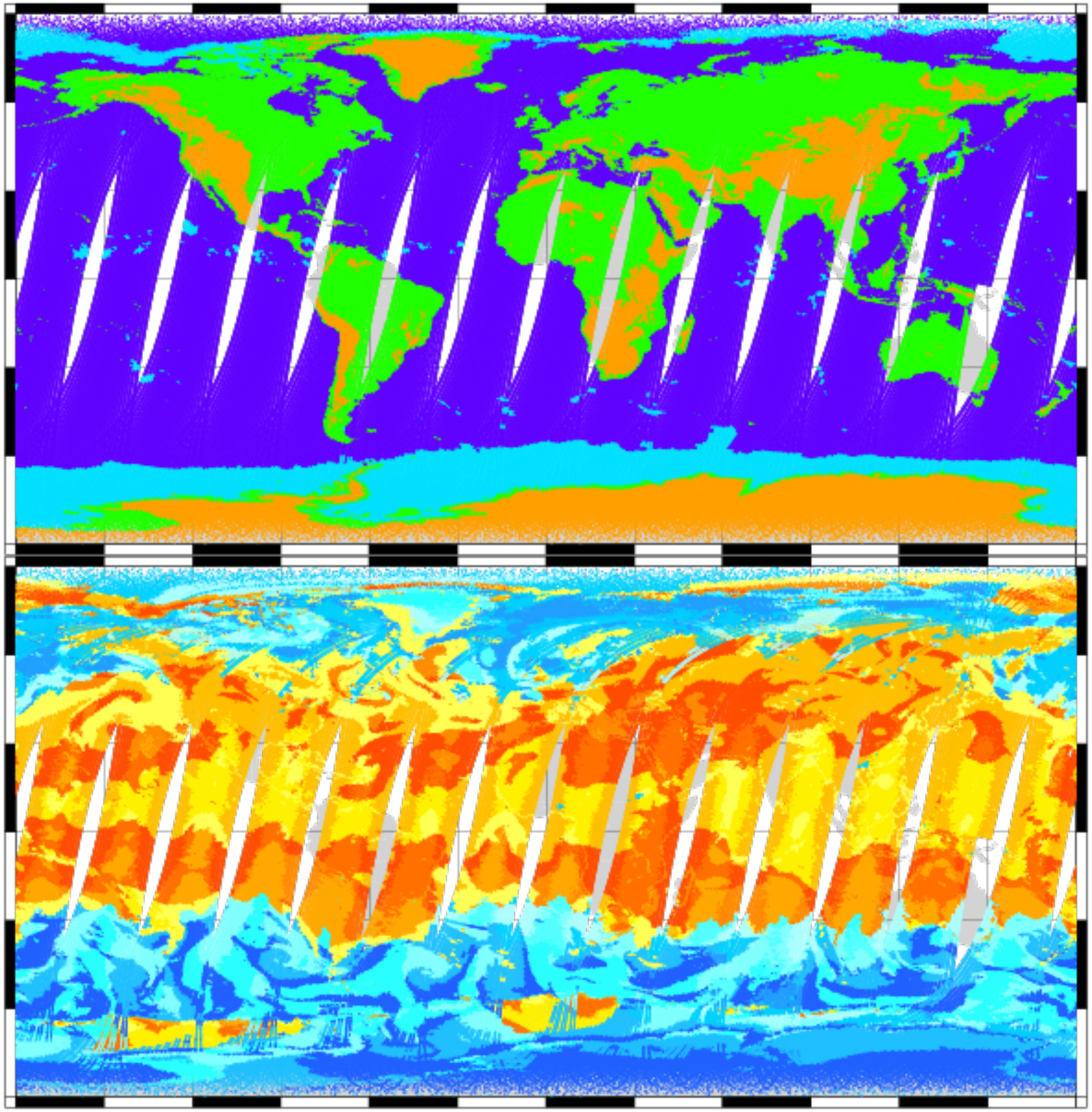
Each class further subdivided in 2 according to AMSU_4

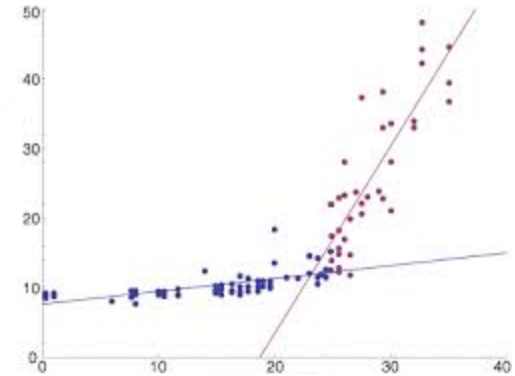
Each class further subdivided in 2 according to MHS_3

Each class further subdivided in 2 according to AMSU_1

Each class further subdivided in 2 according to AMSU_12

Piece-wise linear regression (64 regression classes)



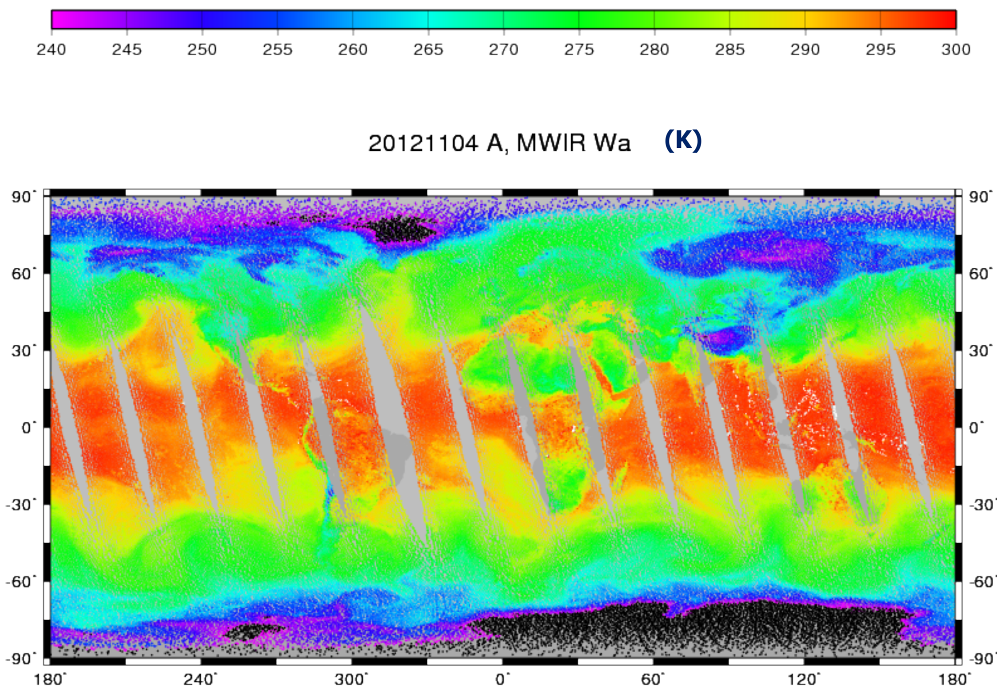


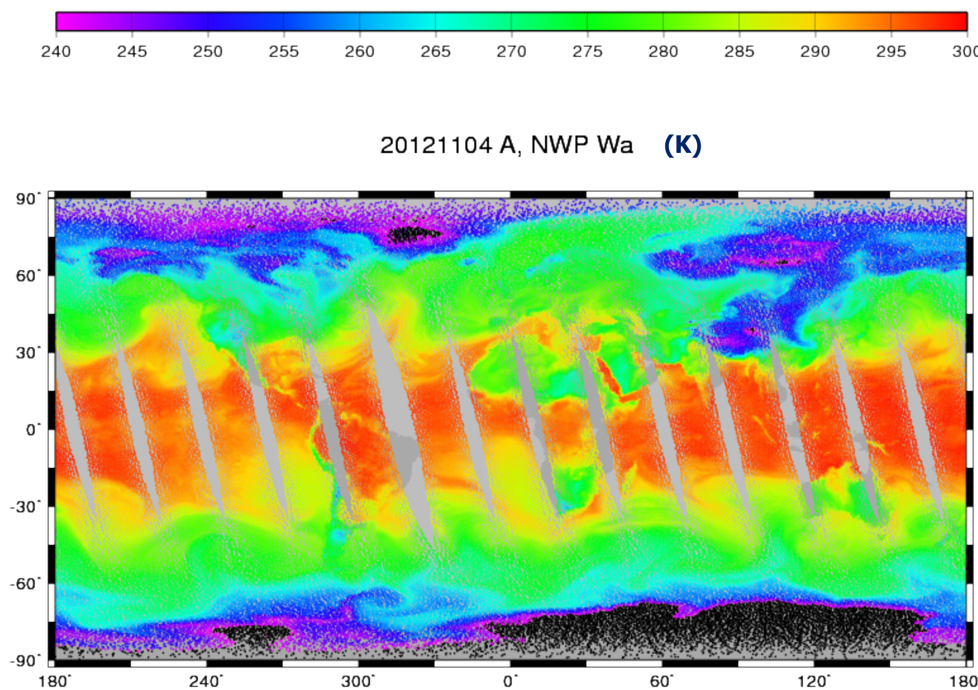
For each class a different set of regression coefficients are used.

Better ability to capture non-linearity in the relation between measurements and state-vector.

Classification depends on measurements only. No explicit dependence on geo-location or time.

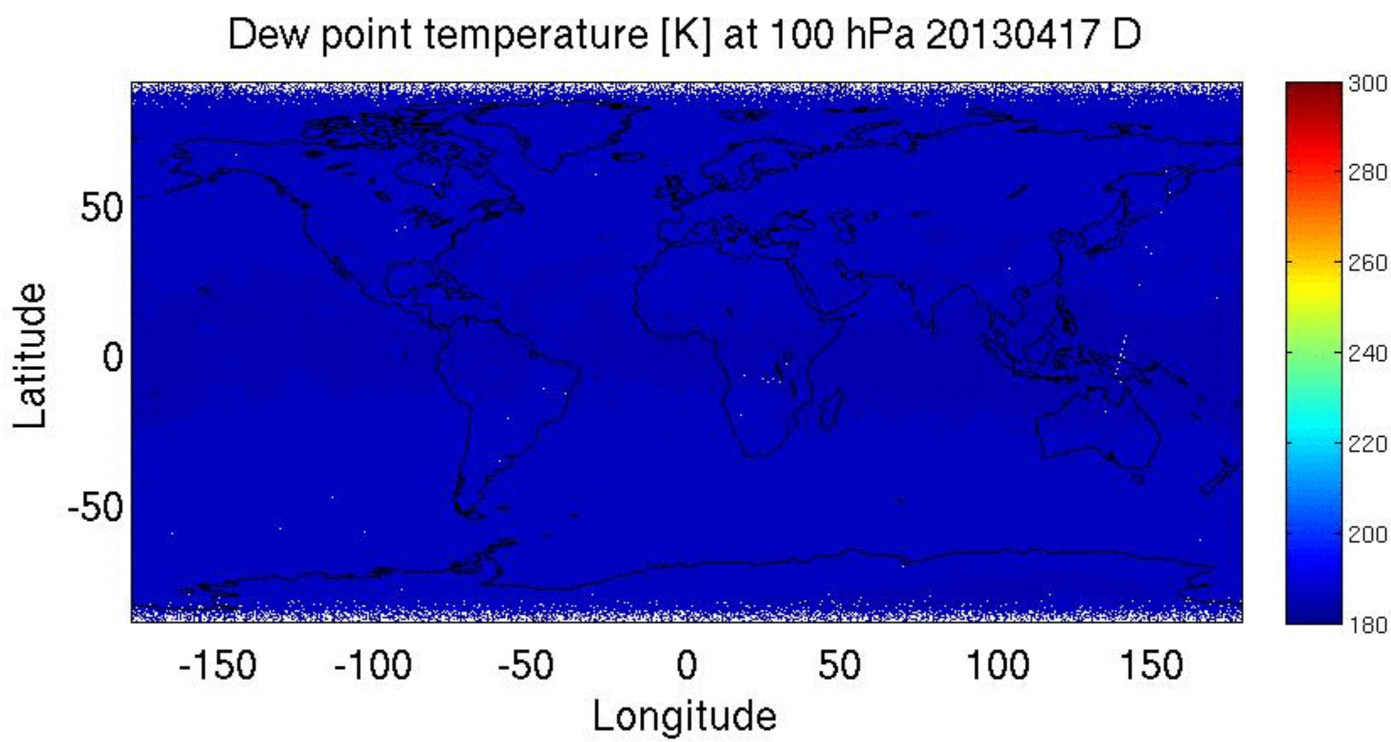
Surface air dew point temperature





EUMETSAT

PWLR retrieved humidity from IASI+AMSU/MHS - 17 April 2013

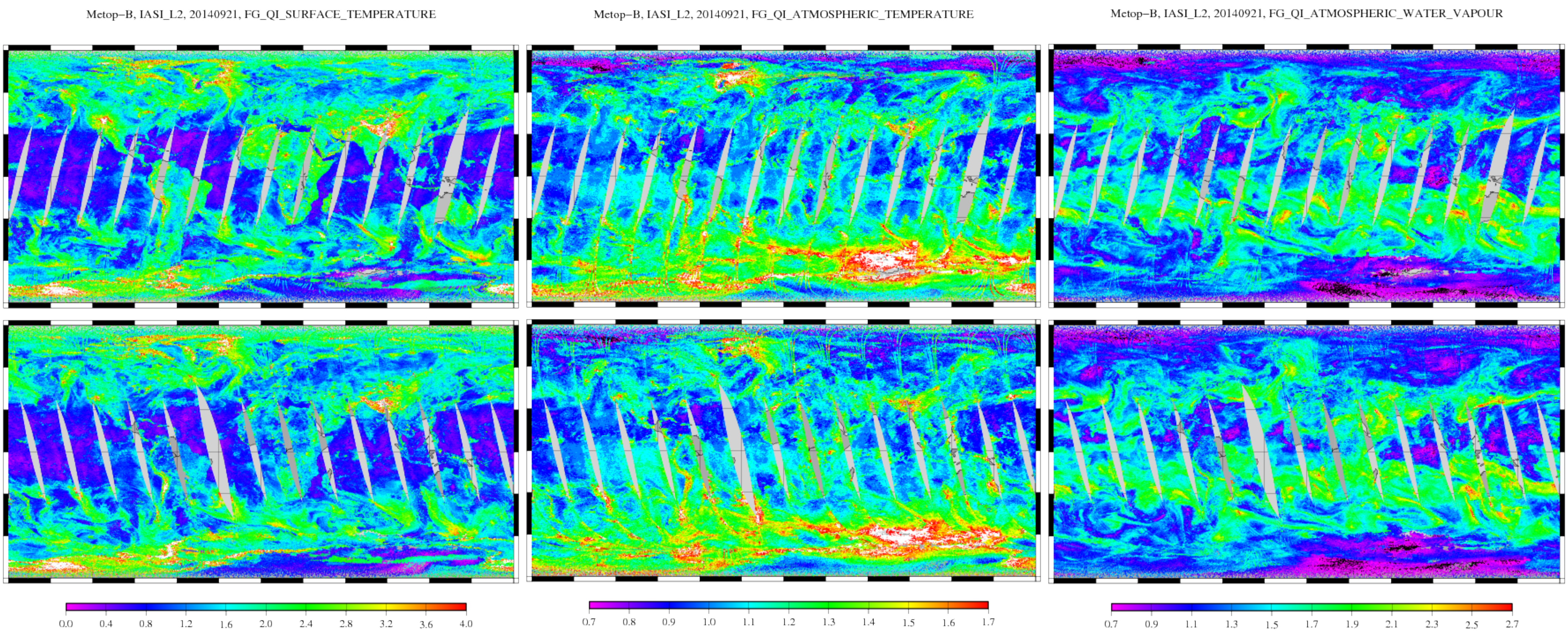


PWLR Quality indicators

Regression coefficients to predict the absolute value of the retrieval error used to derive quality indicators for all retrieved parameters

$$y \approx \bar{y} + R(x - \bar{x})$$

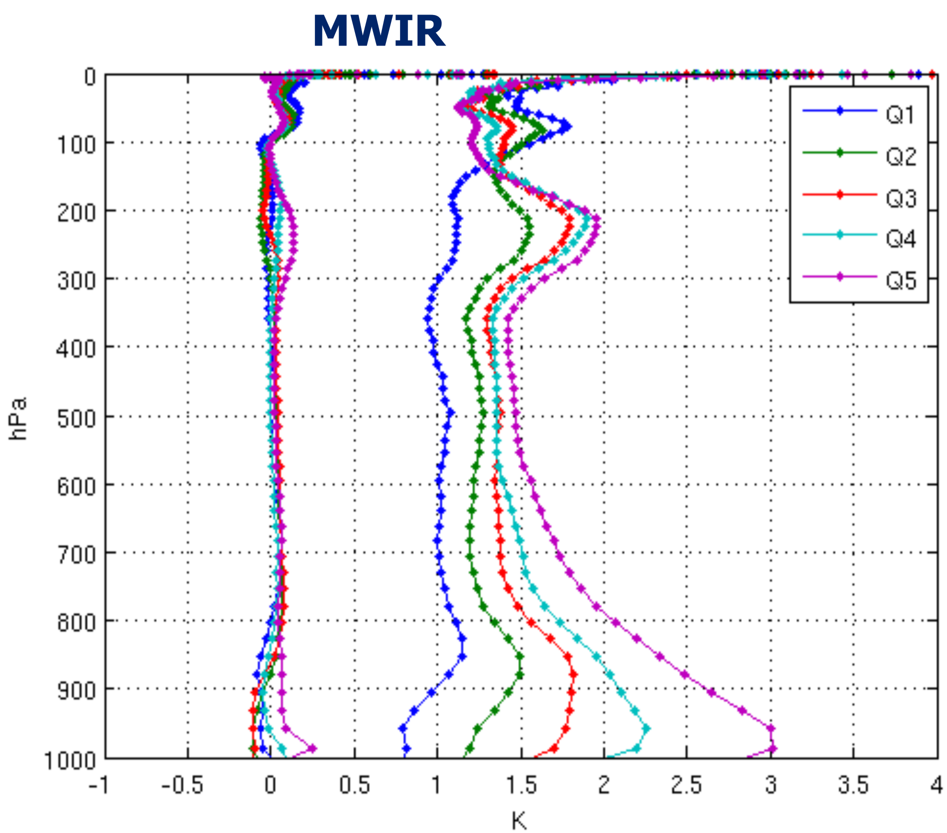
$$\left|\bar{y} + R(x - \bar{x}) - y\right| \approx \overline{\left|\bar{y} + R(x - \bar{x}) - y\right|} + R^{E}(x - \bar{x})$$

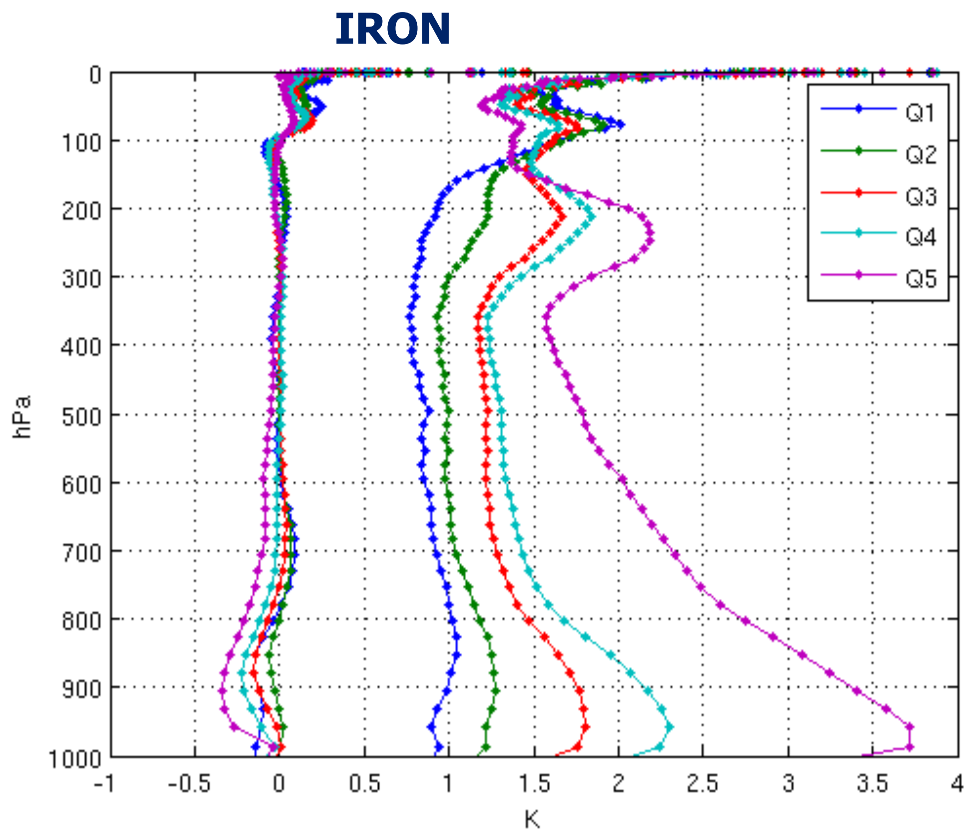


EUMETSAT

Global temperature retrieval minus ECMWF statistics

Quality classes defined by value of the quality indicator for Ta





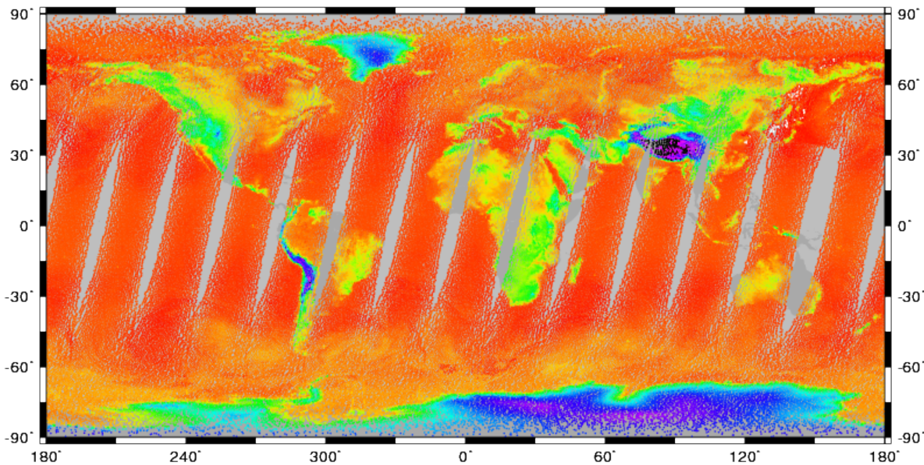
	Q1	Q2	Q3	Q4	Q5
MWIR	15.6%	15.6%	15.6%	15.6%	37.9%
IRON	5.7%	8.6%	7.8%	8.6%	69.3%

EUMETSAT

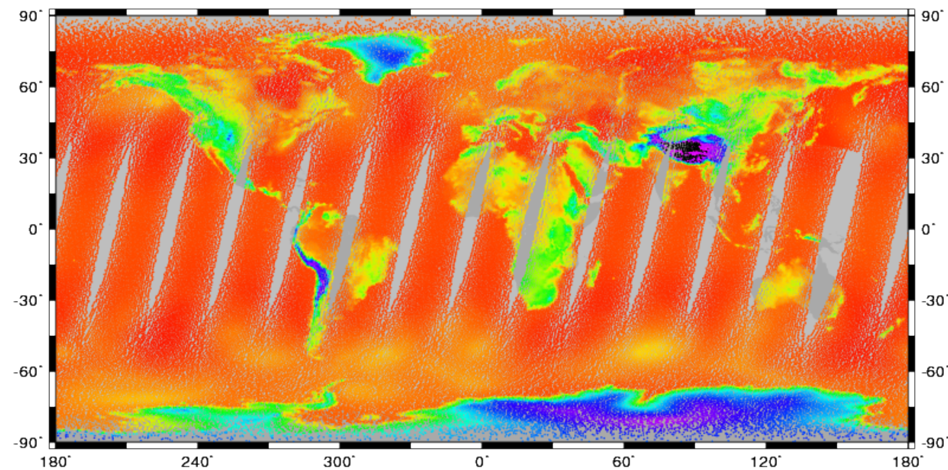
PWLR retrieved surface pressure

Used for interpolation of model levels to fixed pressure grid and as FRTM input in OE

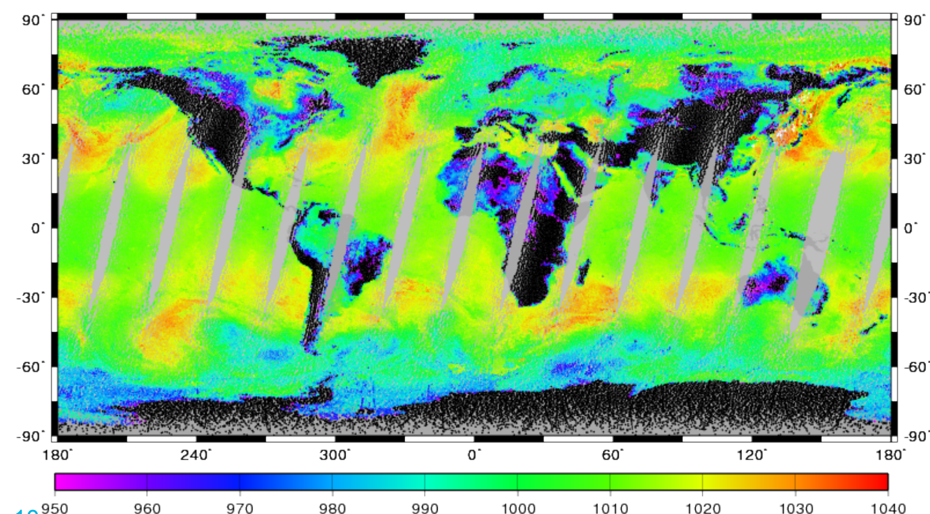
20121104 D, MWIR Sp



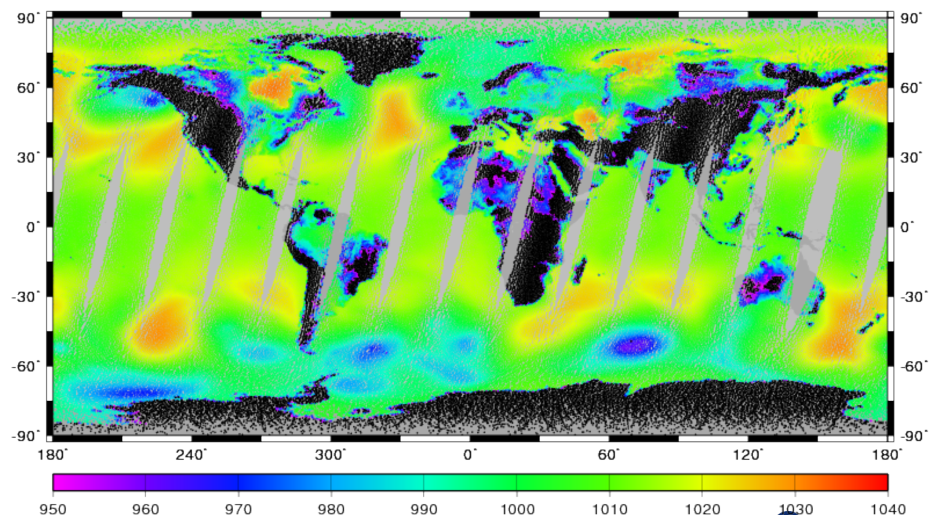
20121104 D, NWP Sp



20121104 D, MWIR Sp



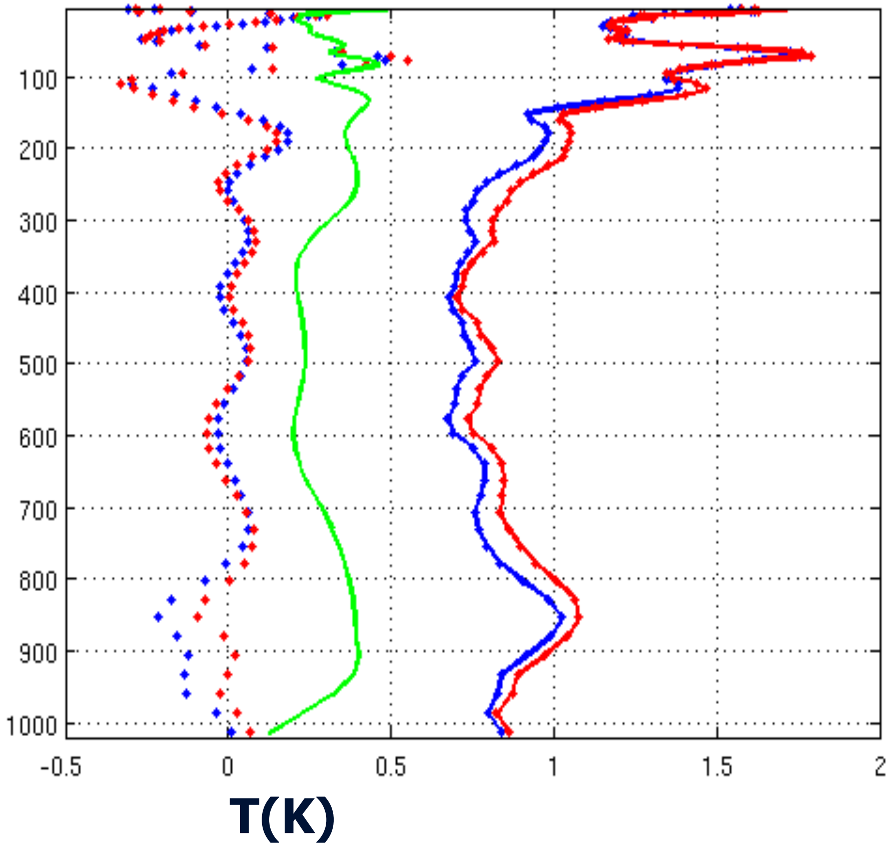
20121104 D, NWP Sp

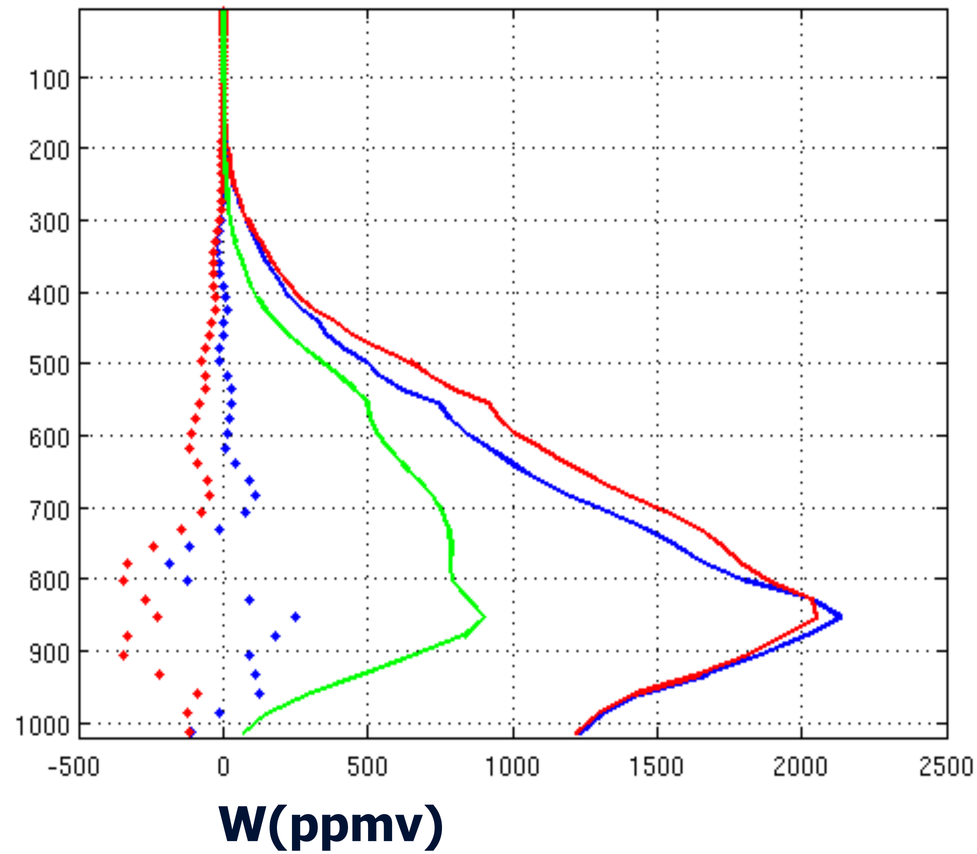


EUMETSAT

Retrieval comparison with ECMWF analysis

20120401 sea only. PWLR - ECMWF, 1DVar - ECMWF, 1DVar - PWLR





Key points of the optimal estimation scheme

Forward model is RTTOV-10.2 (with coefficients trained with LBLRTM v12.2)

State-vector representation:

T: 28 PC scores → 101 (K)
W: 18 PC scores → 101 (log(ppmv))
O: 10 PC scores → 101 (log(ppmv))
Ts: (K)

A priori from PWLR

139 channels (62 in band 1 and 77 in band 2)

Measurement space: Reconstructed radiances (filtered for instrument artefacts)

Full observation error covariance matrix based on OBS-CALC(PWLR)

Object oriented implementation in C++ with abstract interfaces:

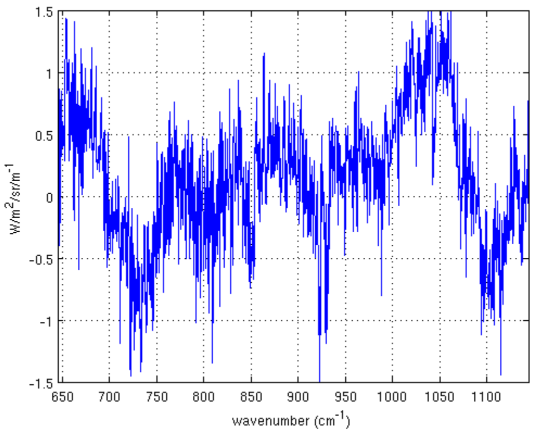
Orthogonal problems implemented by independent classes gives flexibility and facilitates testing and evolution

Reconstructed radiances

Obtained by projection of raw radiances onto the “signal” subspace

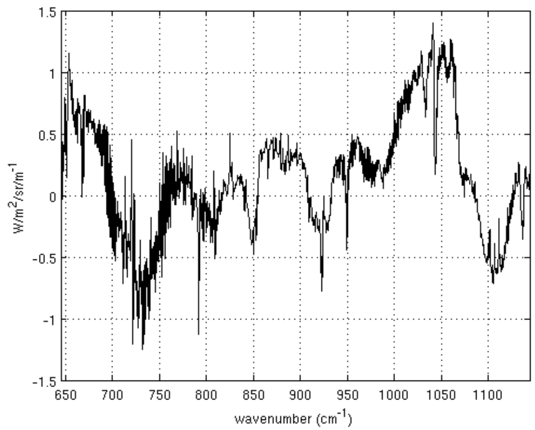
	Total		Signal		Noise
Raw radiance	y	$=$	y_0	$+$	ε
Reconstructed radiance	$\widetilde{y}$	$=$	$\boxed{\widetilde{y}_0}$	$+$	$\boxed{A\varepsilon}$
Residual	$y-\widetilde{y}$	$=$	$\boxed{y_0-\widetilde{y}_0}$	$+$	$\boxed{(I-A)\varepsilon}$

Raw radiance
(minus calc(apriori))



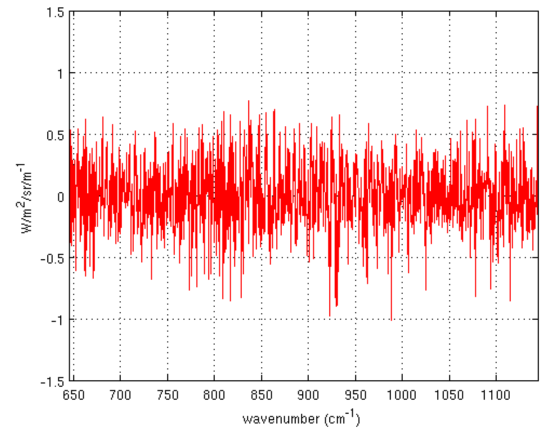
=

Reconstructed radiance
(minus calc(apriori))



+

Residual

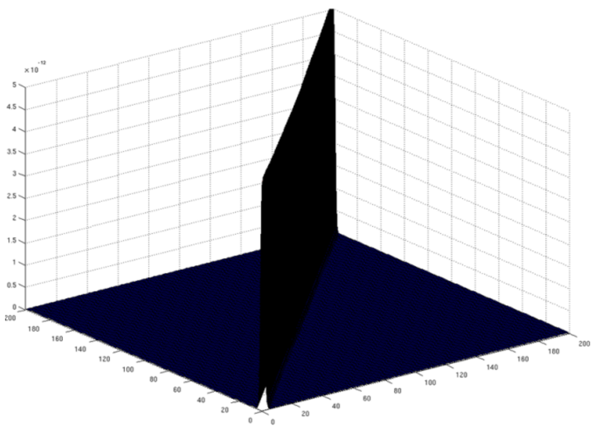


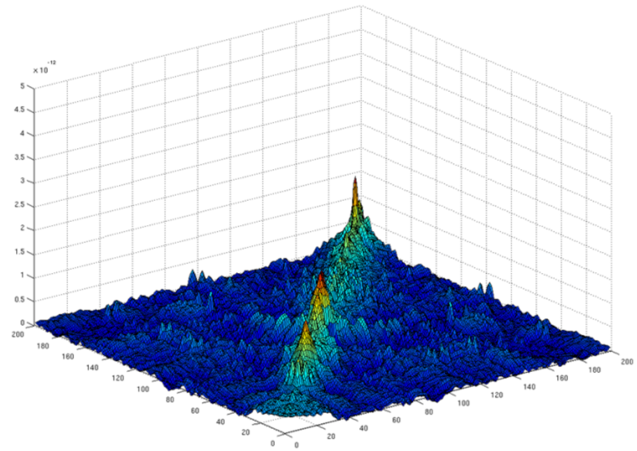
Channel selection (reconstructed radiances)

IASI spectra can be represented by a small number of PC scores with only minor loss of information. The same is true for a small number of reconstructed radiances.

In fact the cost function in both cases are identical if the channel subset is chosen such that the observation error covariance matrix in reconstructed radiance space is non-singular

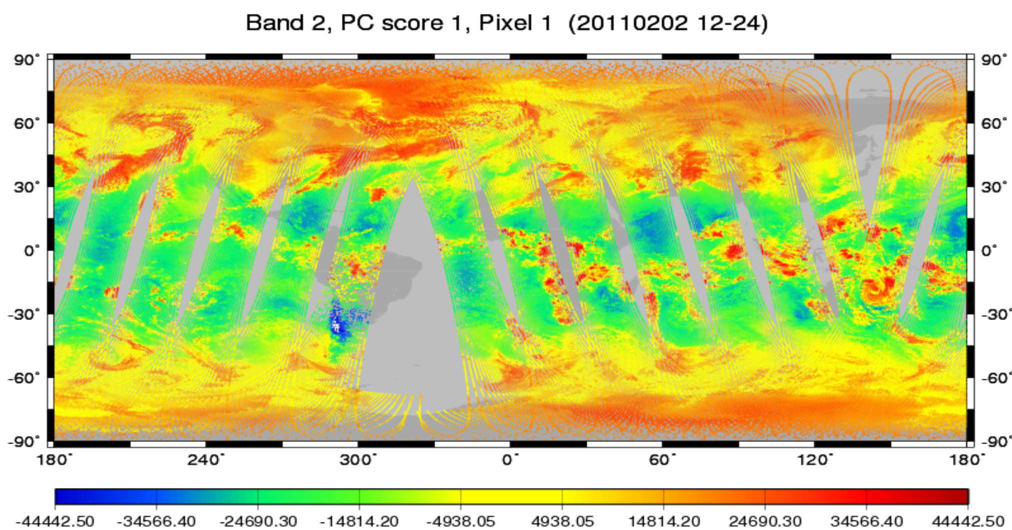
(rank m) S_y → $EE^TS_yEE^T$ (rank s)

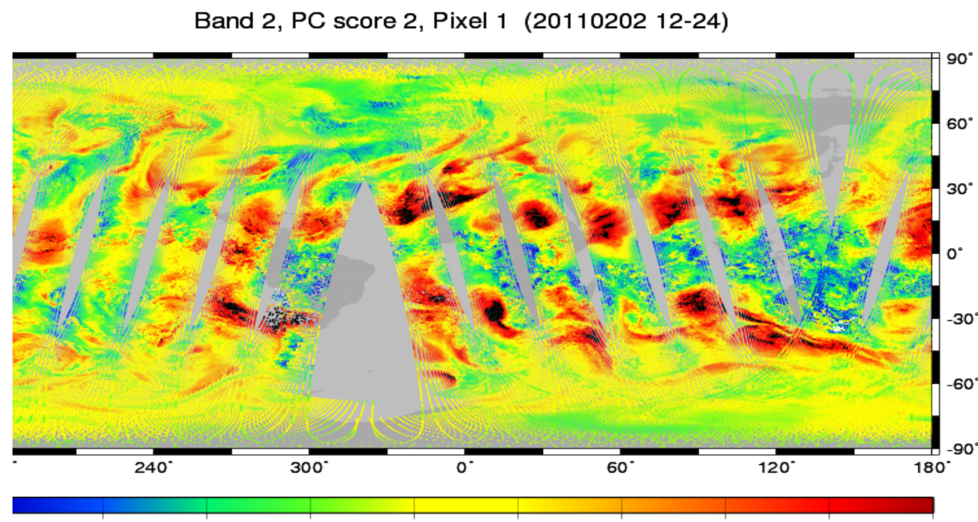


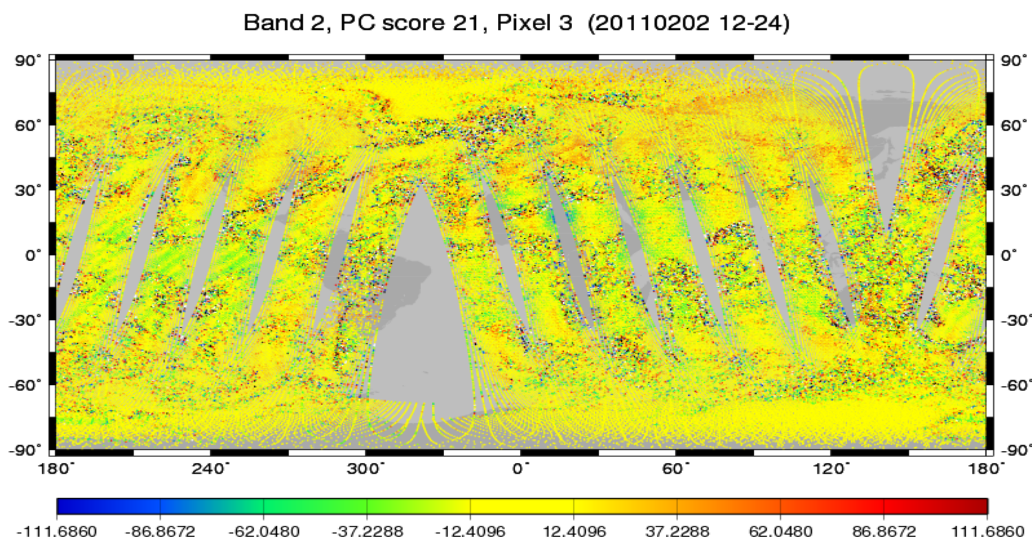


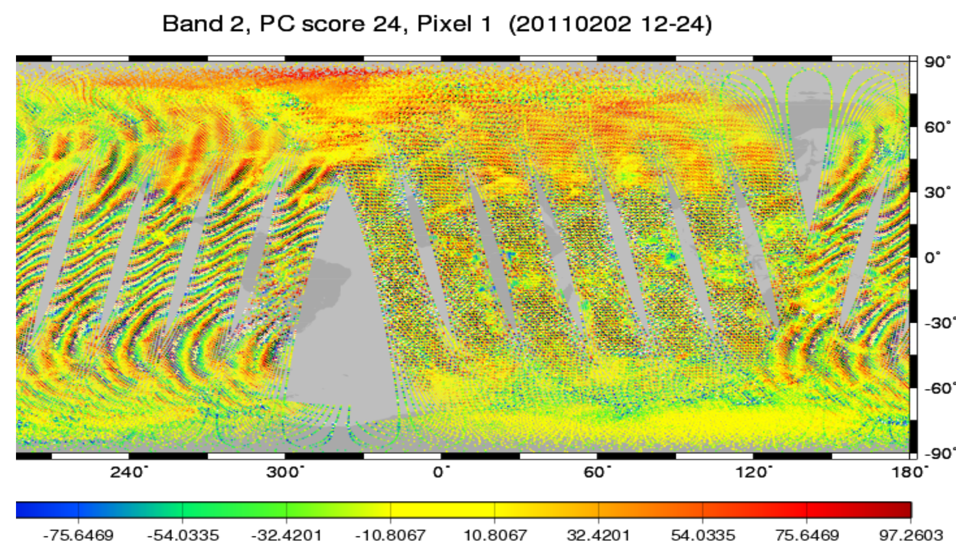
Need to select s channels such that the corresponding sub-matrix of $EE^TS_yEE^T$ is non singular, which is equivalent to selecting s linearly independent rows of the matrix square root $EE^TS_y^{1/2}$ for example by Gaussian elimination. The condition number of the observation error covariance matrix for reconstructed radiances can be minimized (heuristically) by choosing numerically large pivot elements.

Plots of selected PC scores – atmosphere and instrument

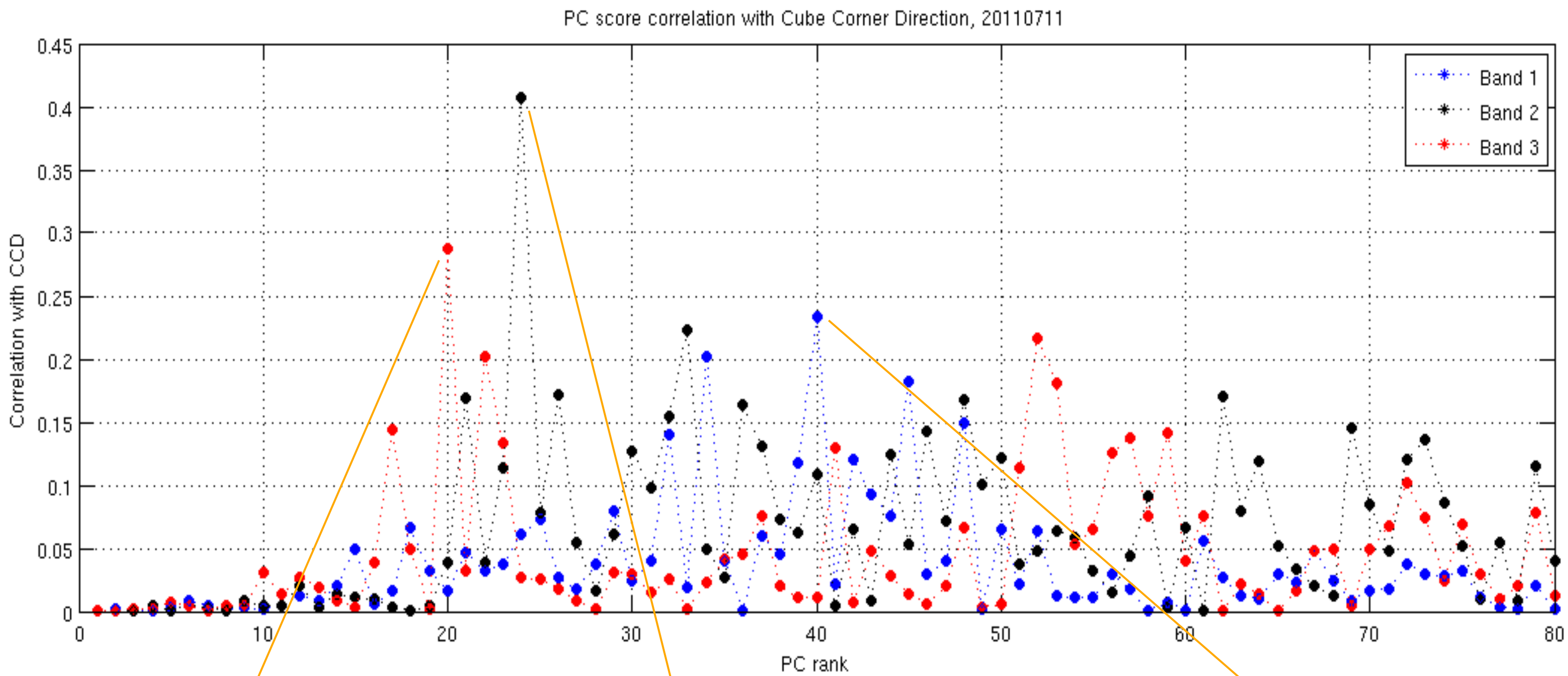


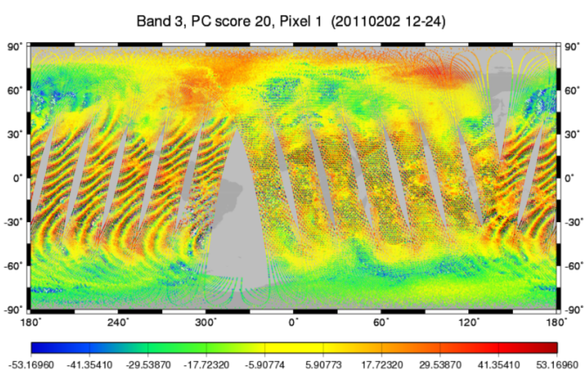


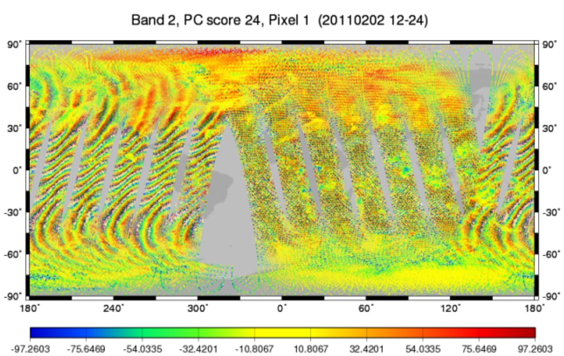


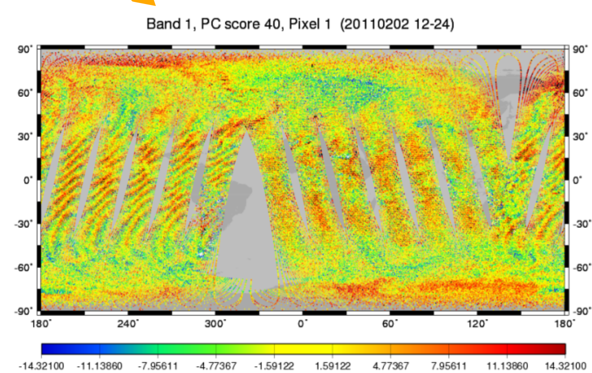


Correlation with cube corner direction



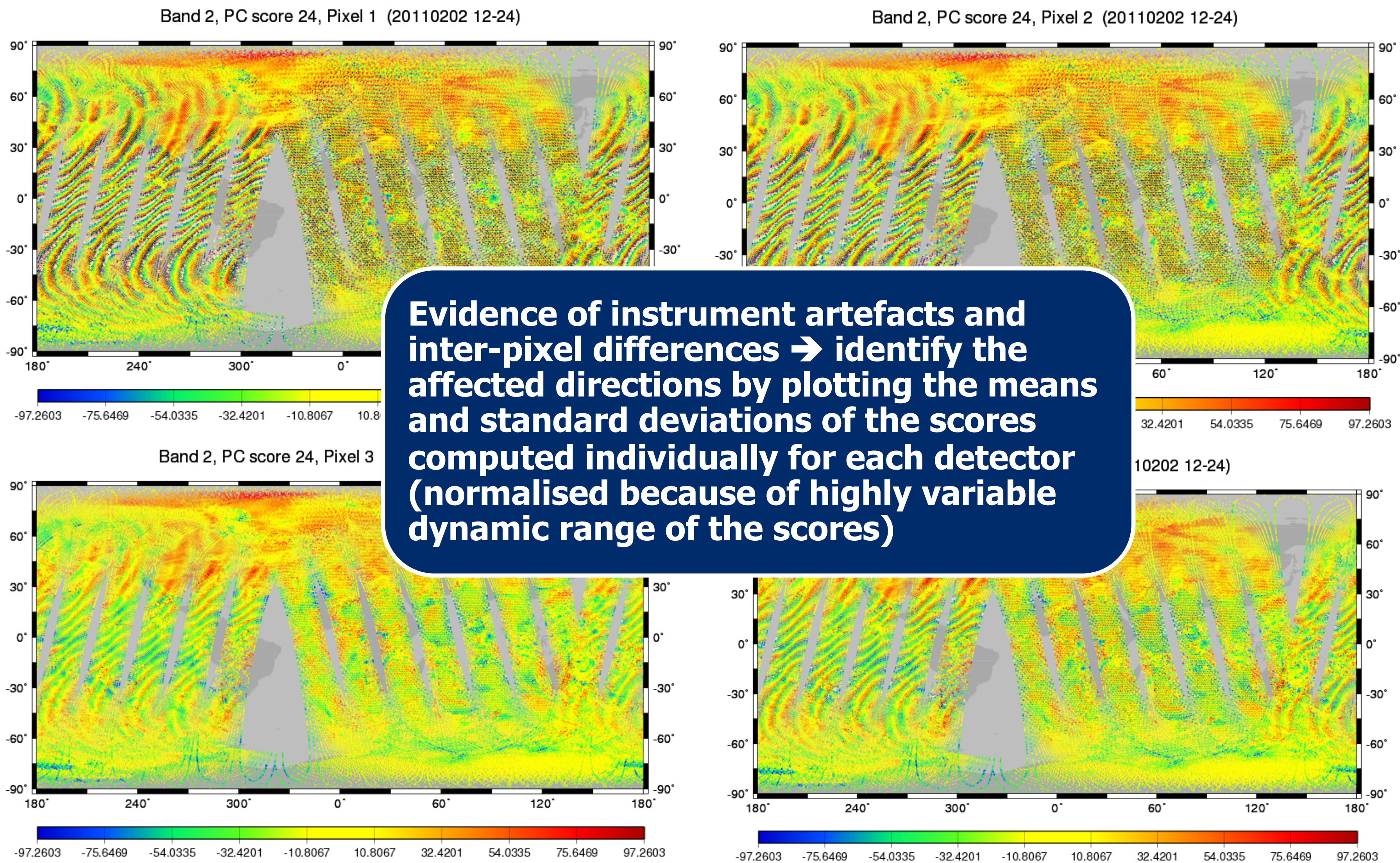






EUMETSAT

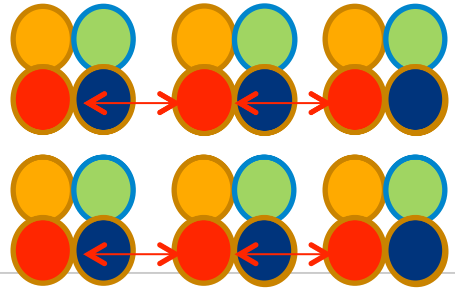
Inter-pixel differences

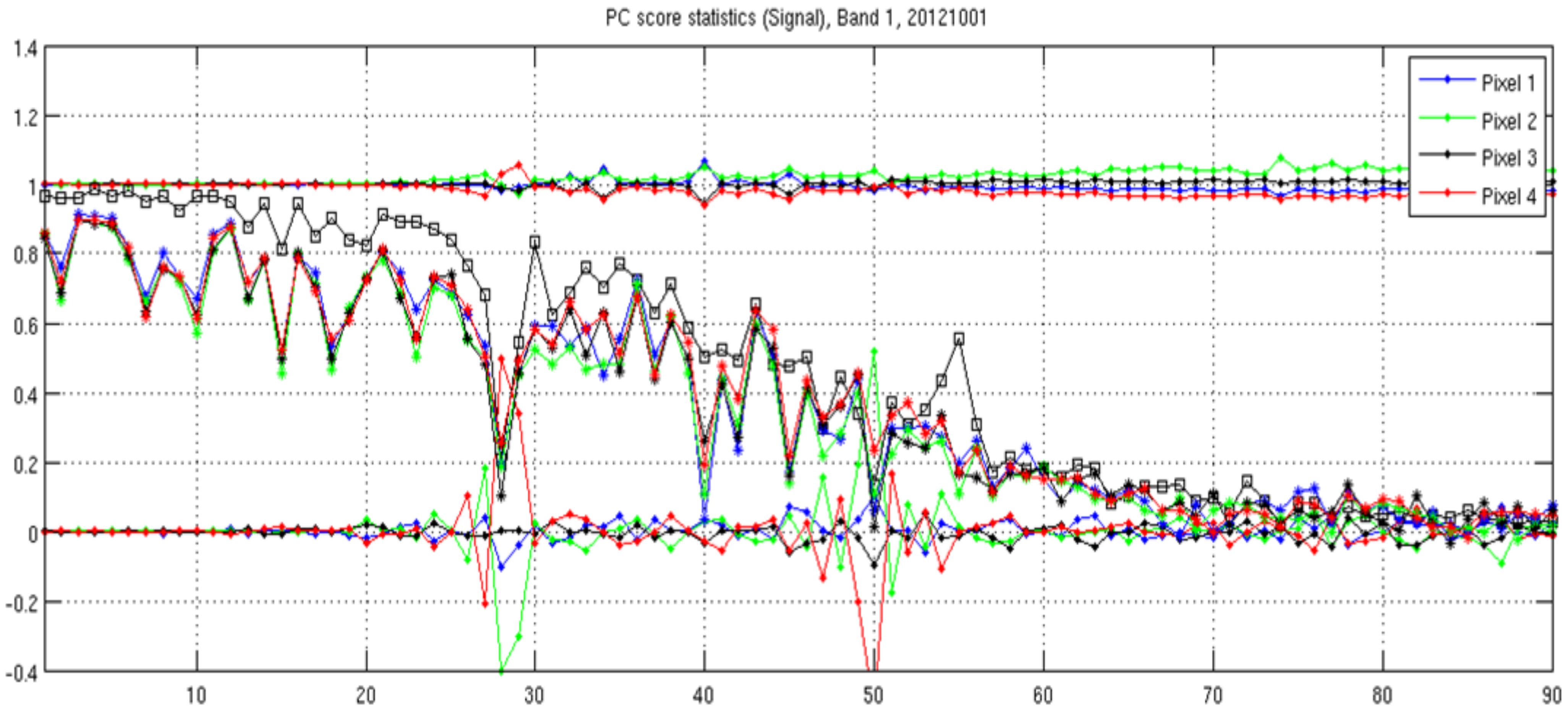


Evidence of instrument artefacts and inter-pixel differences → identify the affected directions by plotting the means and standard deviations of the scores computed individually for each detector (normalised because of highly variable dynamic range of the scores)

PC score statistics for detection of instrument artefacts

- ➢PC score standard deviation (divided by average standard deviation) per detector
- ➢PC score mean (minus average mean, divided by average standard deviation) per detector
- ➢PC score spatial correlation per detector and inter-EFOV

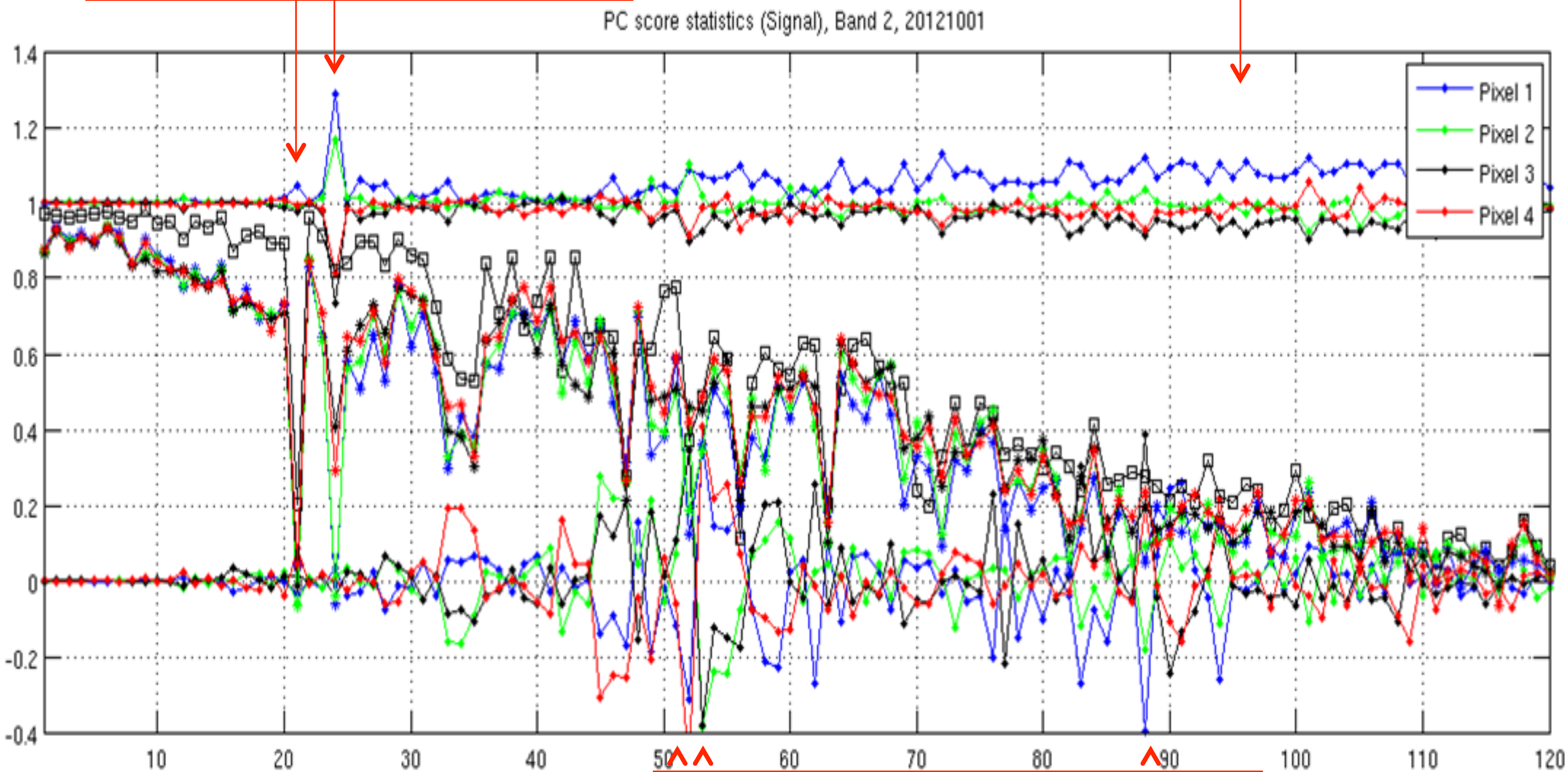




Band 2 PC score statistics

Very low spatial correlation for score 21 and 24

Noise highest in Pixel 1



Several directions with very different means in the 4 pixels

Canonical angles between signal and model subspaces

The subspaces are determined by truncated set of eigenvectors of the covariance matrices of the measured and simulated radiances respectively

$$E_S \in R^{m \times p} \qquad E_F \in R^{m \times p}$$

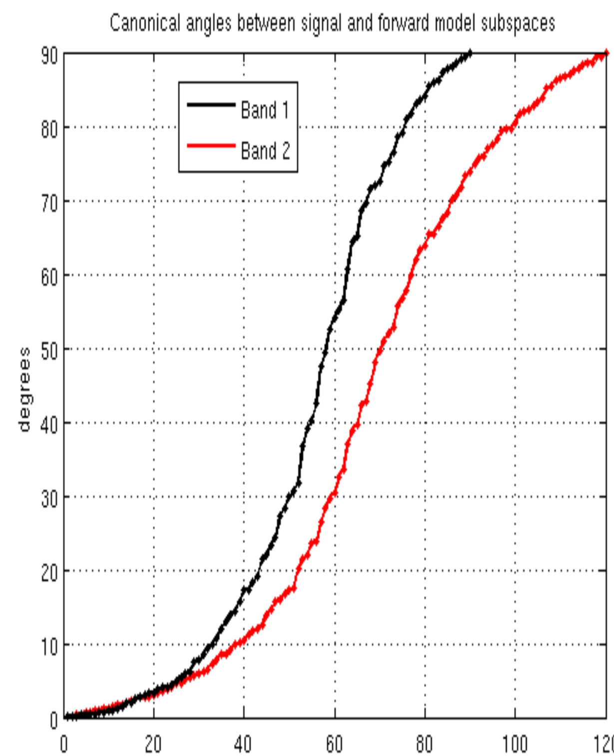
The intersection of the two subspaces is empty. But clearly directions very close to each other can be found in the two subspaces.

$$E_S{}^T E_F = USV^T$$

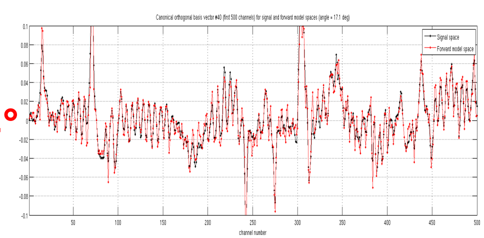
$$\widehat{E_S} = E_S U \qquad \widehat{E_F} = E_F V$$

$\hat{E}_F$ and $\hat{E}_S$ are bi-orthogonal and the canonical angles between the two subspaces are given by arccos(S_{ii}) in ascending order

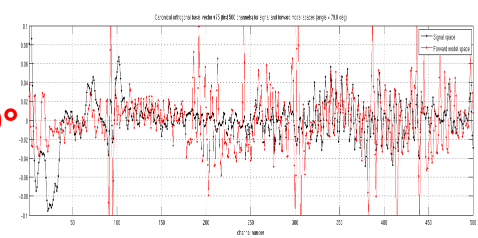
New bases for the signal and forward model spaces, in which similar directions are identified and ordered according to their degree of similarity



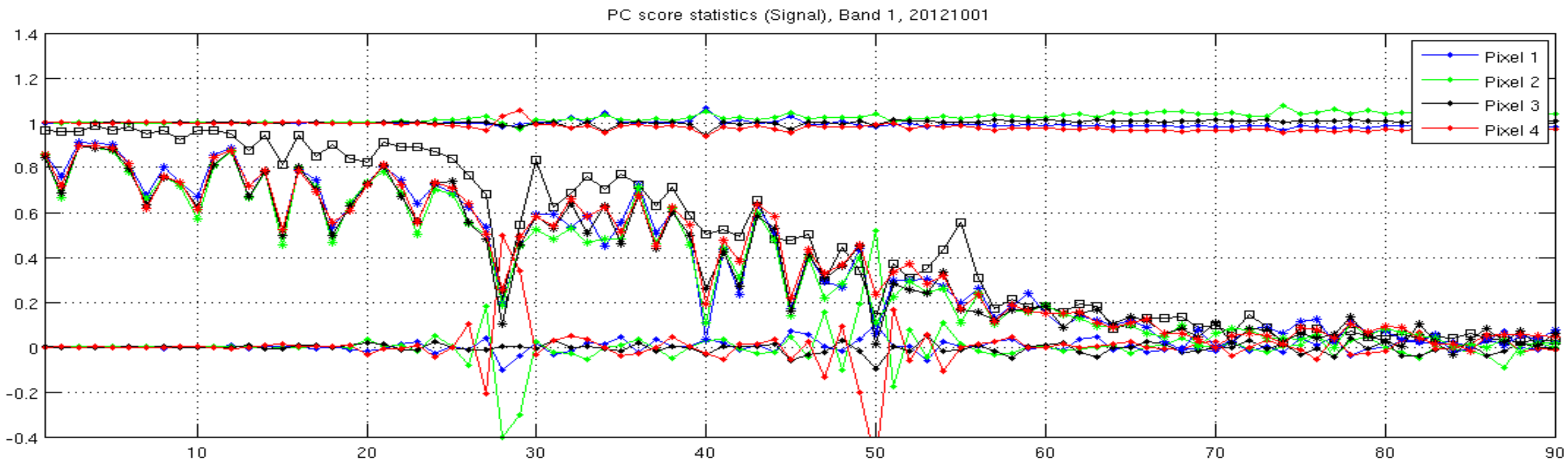
17.1°

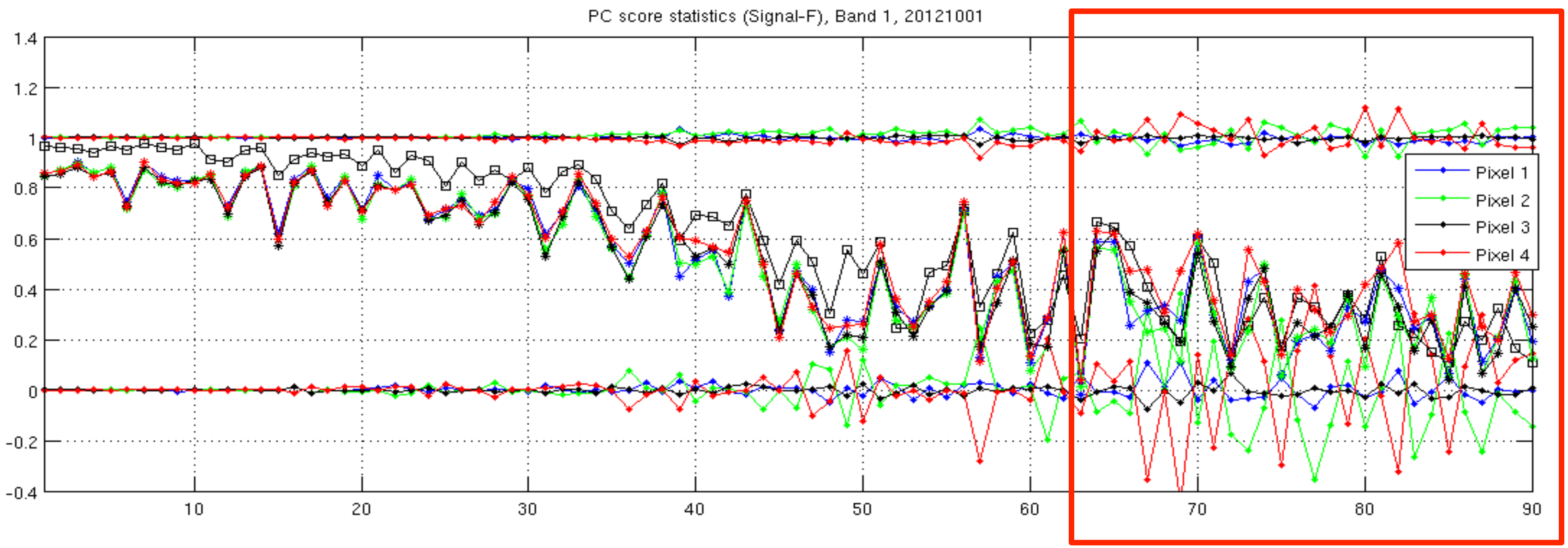


79.0°



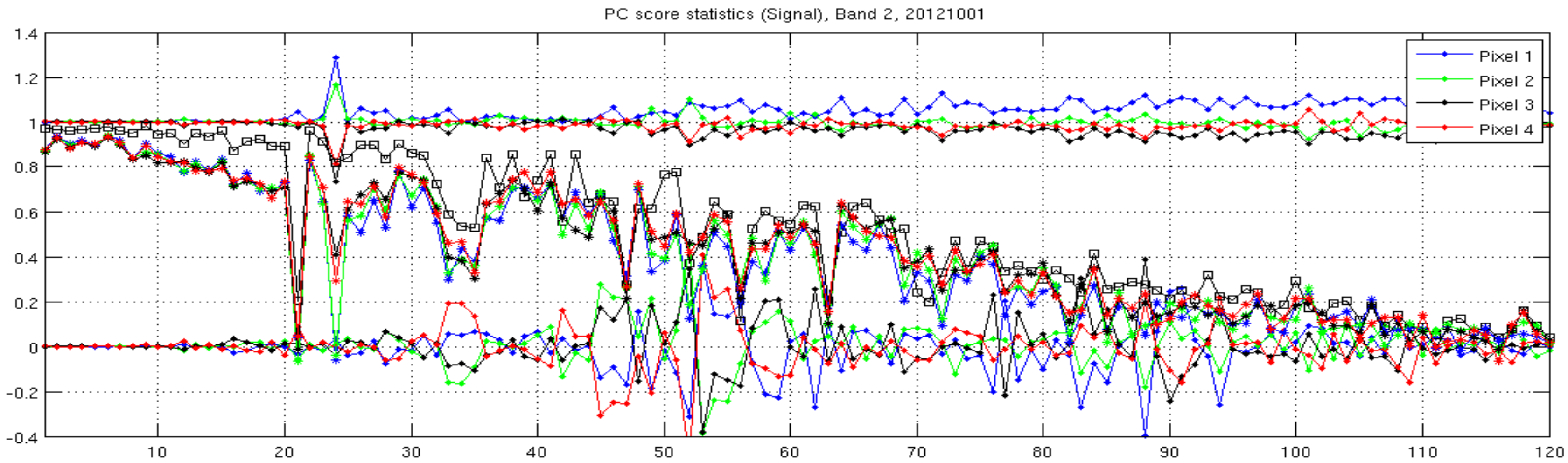
PC score statistics, Band 1

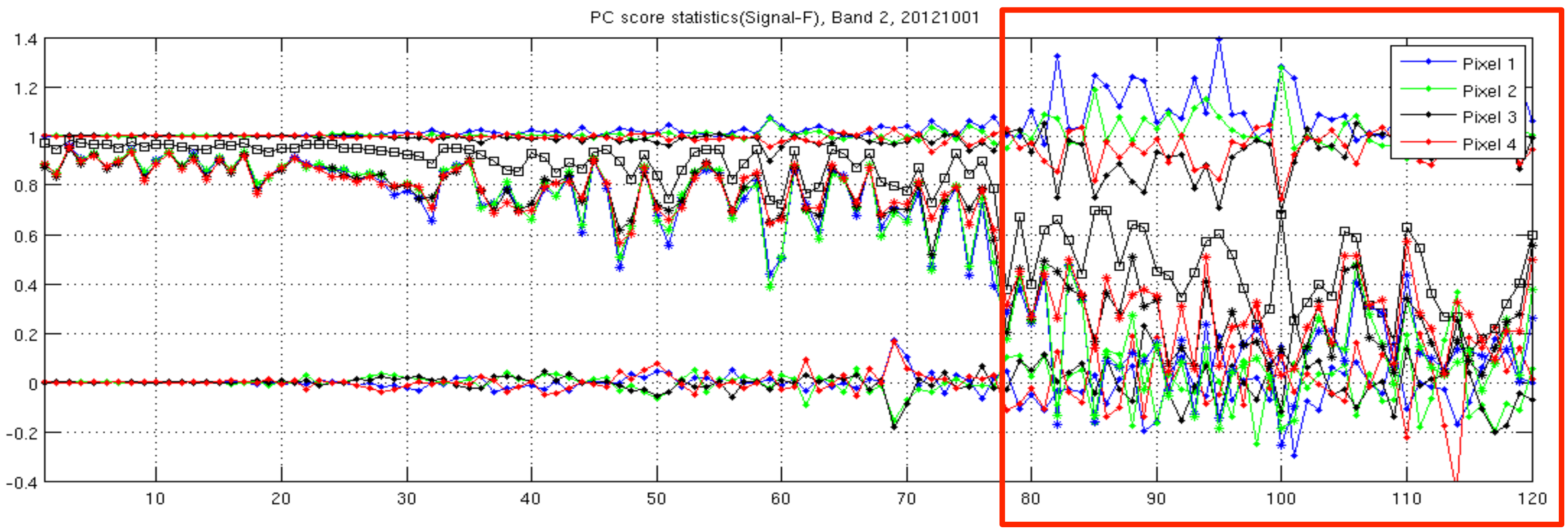




EUMETSAT

PC scores statistics, Band 2

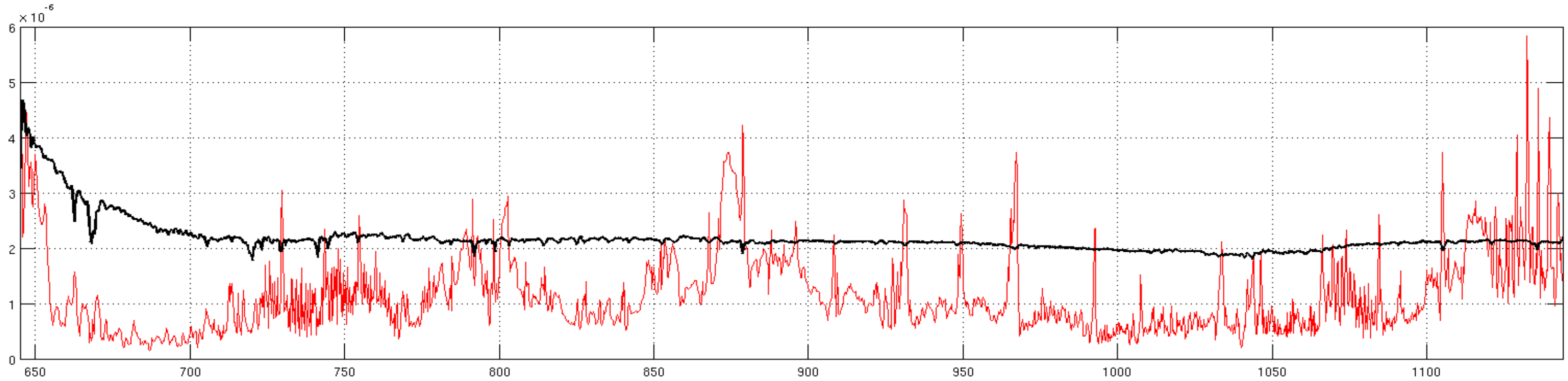




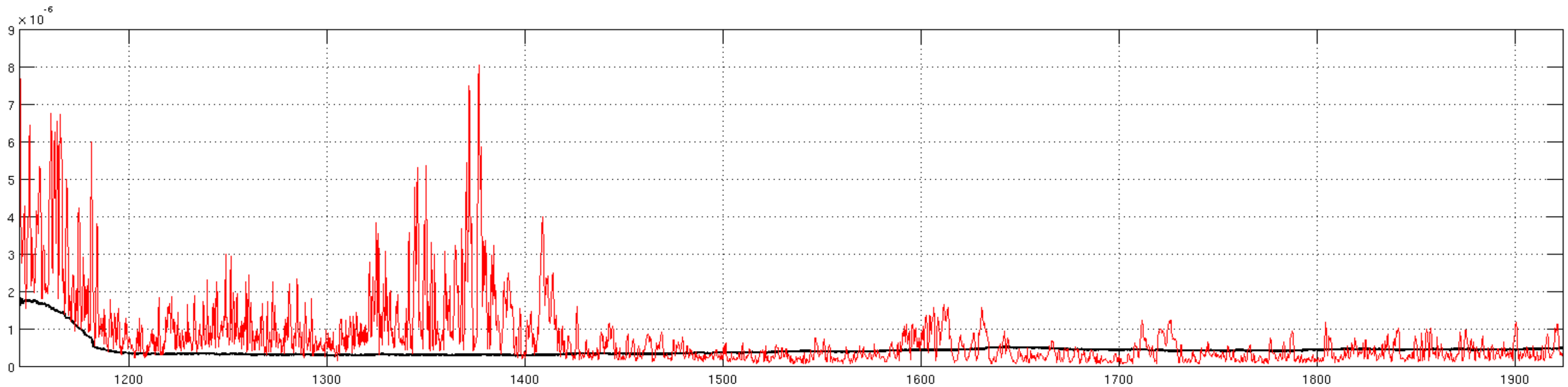
EUMETSAT

Magnitude of instrument artefacts removed by filtering

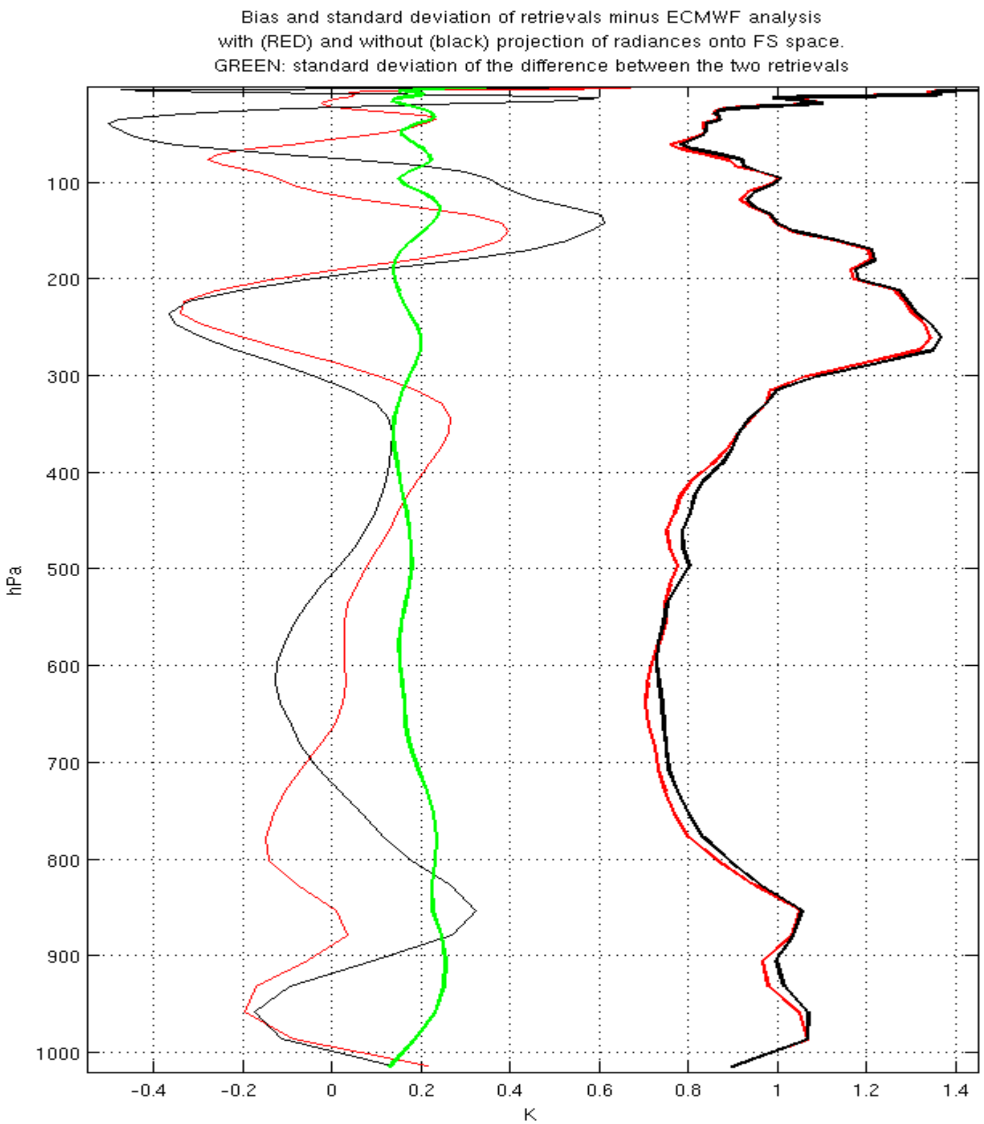
Band 1



Band 2



Impact of filtering onto the FS space



Radiances projected onto signal space

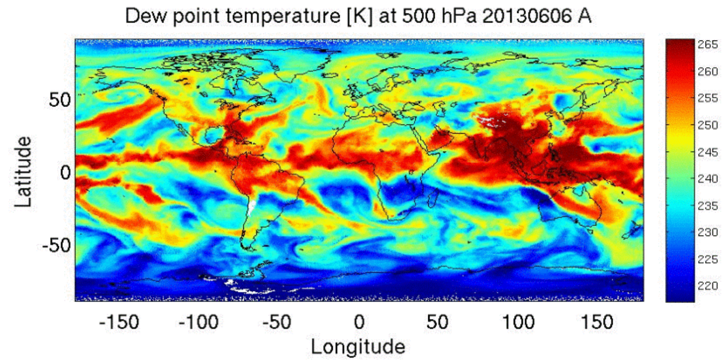
Radiances projected onto the FS space

Standard deviation of the difference of the retrievals with and without projection onto FS space

Statistics based on 11822 cases over sea, +/- 60° latitude on the 2012.10.01

EUMETSAT

Conclusions



PWLR works well!

- **Synergistic use of MW**
- **Comes with reliable quality indicators**
- **Trained with real measurements → Handles features not modelled by the RTM**
- **Trained with BIG datasets → Insensitive to random errors in reference data**
- **Easy and efficient way to handle non linear response**
- **Evidence of scope for improvement (to be introduced in due time)**

But any systematic biases in reference data (ECMWF analysis) are retained

Optimal estimation

- **Usually converges after 1 or 2 full Newton steps**
- **Reconstructed radiances and full matrix observation error covariance**
- **Forward model only invoked for 139 channels**
- **Configuration tuned against ECMWF, but retrievals do not use forecast**
- **Improved by removal of instrument artefacts from the measurements**
- **Improved by use of good a-priori (PWLR)**